

# Agentes inteligentes en ambientes dinámicos: Simplificación de Texto

Enzo Ferrarini

Sandra Roger

email: {ferrarini.enzo.nqn,giuvago}@gmail.com

*Grupo de Investigación en Lenguajes e Inteligencia Artificial*

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300) Neuquén - Argentina

## Resumen

La meta fundamental de este proyecto está puesto tanto en los formalismos de planificación y razonamiento rebatible para la creación y control de agentes inteligentes, como en el impacto que tienen las tecnologías del lenguaje humano (TLH) en la inclusión social. En estos escenarios, el razonamiento, la toma de decisiones, la planificación de acciones y el aprendizaje ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario.

Una de las líneas de investigación es la simplificación de texto. El objetivo de esta es el desarrollo de tecnologías para facilitar el acceso a la información para todas aquellas en el que la inteligibilidad del texto es de gran importancia.

**Palabras Clave:** AGENTES INTELIGENTES, LENGUAJE NATURAL, SIMPLIFICACIÓN DE TEXTO, ACCESIBILIDAD.

## Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación *Agentes inteligentes en ambientes dinámicos*. El

proyecto de investigación tiene prevista una duración de cuatro años, desde enero del 2013 hasta diciembre de 2016.

## 1. Introducción

La simplificación de texto es una metodología utilizada en muchos idiomas para conocer las necesidades de los lectores que tienen diferentes niveles educativos y de capacidad de comprensión. Además proporciona una solución para los creadores de contenidos interesados en mejorar la sencillez y, por tanto, la accesibilidad de su contenido.

La forma en la que están estructuradas las sentencias, sean estas por su longitud y complejidad, ocasionan un gran problema tanto a nivel de comprensión como a nivel de entendimientos del texto en sí mismo. Como se menciona en [2]: “Si la complejidad del texto puede hacerse más simple, las oraciones resultarán más fáciles de procesar tanto para las personas como para los programas”.

Las características claves en el proceso de comprensión de lectura de un texto son: el texto, el lector y las circunstancias en que se celebra la reunión. El análisis de estudios publicados muestra que la comprensión lectora implica varios factores que se pueden dividir en tres

grupos: i) en el texto, ii) para el lector, y iii) en relación con intervención pedagógica. Entre los factores relacionados con el texto se destacan, tradicionalmente, la legibilidad (formato del texto) y la inteligibilidad (uso frecuente de palabras y estructuras sintácticas menos complejas). Es bien sabido que las sentencias largas, con varios niveles de subordinación, cláusulas incrustadas (relativa), las frases en voz pasiva, el uso de no-canónico para que los componentes de una frase, y el uso de palabras de baja frecuencia aumentan la complejidad de un texto para lectores con problemas de comprensión, por ejemplo, analfabetos funcionales, la afasia y la dislexia [6].

En la actualidad, también hay una preocupación con la macroestructura del texto más allá de la microestructura. En esta última se observan otros factores, los cuales son vistos como facilitadores en la comprensión de texto, tal como la organización, la cohesión, la coherencia, o conceptos del texto sensibles al lector. Este último presenta características que pueden facilitar la comprensión como aproximación de la anáfora, el uso de marcadores discursivos entre las oraciones o una preferencia por definiciones explícitas [5].

En este trabajo se pone principal énfasis, en primera instancia, en la recopilación organizada y metodológica que provee la revisión sistemática sobre la simplificación de texto. Posteriormente se trabajará sobre las implementaciones necesarias que se desprendan de la revisión descripta anteriormente.

## 2. Líneas de investigación y desarrollo

El proyecto de investigación *Agentes inteligentes en ambientes dinámicos* tiene varios objetivos generales. Por un lado, el de desarrollar conocimiento especializado en el área de Inteligencia Artificial. Además, se estudian técnicas de representación de conocimiento y razonamiento, junto con métodos de planificación y tecnologías del lenguaje natural aplicadas al desarrollo de sistemas multiagentes.

La producción de textos con diferentes longitudes y complejidades pueden ser abordadas por la tarea de adaptación del texto (TA), una práctica muy conocida en los ambientes educativos. Young [7] y Burstein [1] mencionan dos técnicas diferentes para TA: Simplificación de texto y elaboración del texto. La primera puede ser definida como cualquier tarea que reduce la complejidad léxica o sintáctica de un texto al intentar preservar el significado e información. Esta se puede subdividir en simplificación sintáctica, simplificación léxica, resumen automático, y otras técnicas. En cuanto a la elaboración del texto, su objetivo es aclarar y explicar la información al hacer conexiones explícitas en un texto, por ejemplo, proporcionando definiciones cortas o sinónimos de palabras conocidas a sólo unos pocos hablantes de una lengua.

### 2.1. Revisión sistemática

La revisión sistemática comprende una metodología específica y ordenada con la cual es llevada a cabo un estudio. No es lo mismo hablar de una revisión sistemática que de una revisión clásica o narrativa. La diferencia principal entre ambas, es que en esta última, existe la falta de un protocolo que defina los pasos específicos que utilizó el revisor para actualizar el tópico en cuestión. Por ejemplo, estrategias utilizadas para la búsqueda y selección de la bibliografía y métodos para determinar la validez de los mismos. Se puede definir una revisión sistemática como un resumen de la evidencia existente en una determinada fecha usando una metodología bien definida que intenta disminuir, en lo posible, el sesgo de las revisiones narrativas.

La revisión sistemática se obtiene a través de una metodología compuesta de ocho pasos bien definidos [3, 4]. A partir de los cuales, se obtienen revisiones bibliográficas independientes a diferencia de otros sistemas no metodológicos.

Pasos:

1. Propósito de la Revisión de la Literatura: El primer paso en cualquier revisión requiere que el revisor identifique claramente

te el propósito y objetivos previstos de la misma. El propósito debe ser explícito hacia sus lectores.

2. Protocolo y Formación: Toda revisión que implique a más de un crítico, debe fundamentar el procedimiento detallado a seguir. Esto requiere de un documento escrito, el protocolo detallado y la formación para todos los revisores para asegurar la consistencia en la ejecución de la revisión.
3. Búsqueda de la Literatura: El revisor debe ser explícito en la descripción de los detalles de la búsqueda en la literatura, y las necesidades de explicar y justificar la misma.
4. Pantalla Práctica: también conocido como screening para la inclusión; este paso requiere que el revisor sea explícito acerca de los estudios que fueron considerados para la revisión, y cuales fueron eliminados o descartados sin más examen.
5. Evaluación de la calidad: este paso es también conocido como la detección de la exclusión, el revisor debe de forma explícita exponer los criterios para juzgar que artículos son de calidad suficiente para ser incluido en la síntesis de la revisión. Todos los artículos incluidos necesitan ser marcados por su calidad, dependiendo de las metodologías de investigación empleadas por los artículos.
6. Extracción de datos: Después de que todos los estudios que se deben incluir en la revisión han sido identificados, los revisores deben extraer de forma sistemática la información correspondiente de cada estudio.
7. Síntesis de los estudios: Este paso es conocido también como análisis; implica la combinación de los hechos extraídos de los estudios que utilicen técnicas apropiadas, ya sea cualitativa, cuantitativa o ambas.
8. Redacción de la revisión: El proceso de revisión sistemática de la literatura tiene q

ser reportado detalladamente para que los resultados de la revisión puedan reproducirse de forma independiente.

### 3. Resultados esperados

Uno de los objetivos de esta línea es poder mostrar que es posible realizar una revisión sistemática en el área. La revisión sistematizada se han realizado en áreas como la medicina por ejemplo. Existe evidencia que muestran que en área de la computación, esta metodología solo es aplicada al campo de la ingeniería del software. Se pretende mostrar que es posible aplicarla en esta línea de investigación.

Por otro lado, partiendo de los resultados obtenidos a partir de la revisión, se pretende desarrollar herramientas que faciliten el tema de la inclusión en el ámbito de la simplificación de texto para el idioma español. Actualmente, se está trabajando en la primer etapa de la revisión.

### 4. Formación de Recursos Humanos

Durante la realización de este sistema se espera lograr, como mínimo, la culminación de 2 tesis de grado dirigidas y/o co-dirigidas por los integrantes del proyecto y una tesis de maestría.

Finalmente, es constante la búsqueda de nuevos investigadores como también, la consolidación de investigadores dentro del grupo.

### Referencias

- [1] J. Burstein. Opportunities for natural language processing research in education. In *Proceedings of CICLing*, pages 6–27, 2009.
- [2] R. Chandrasekar, C. Doran, and B. Srinivas. Motivations and methods for text simplification. In *In Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96)*, pages 1041–1044, 1996.

- [3] B. Kitchenham. Kitchenham, 2004 procedures for performing systematic reviews, 2004.
- [4] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering no. ebse 2007-001., 2007.
- [5] C. E. Scarton and S. M. Aluísio. Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o português. *linguamatica*, 2(1):45–61, Abril 2010.
- [6] A. Siddharthan. Preserving discourse structure when simplifying text. In *In Proceedings of the 2003 European Natural Language Generation Workshop*, pages 103–110, 2003.
- [7] D. J. Young. Linguistic simplification of sl reading material: effective instructional practice. In *The Modern Language Journal* 83(3), pages 350–366, 1999.